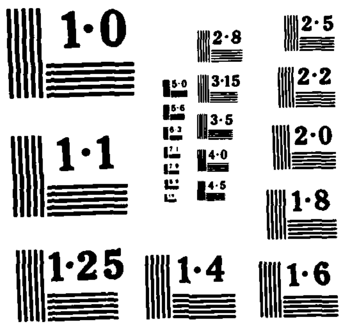
AD-A155 855 CHARACTERIZATIONS OF GENERALIZED HYPEREXPONENTIAL DISTRIBUTIONS(U) VIRGINIA UNIV CHARLOTTESVILLE DEPT OF SYSTEMS ENGINEERING R F BOTTA ET AL. MAY 85 1/1

UNCLASSIFIED UVA/525393/SE85/107 N00014-83-K-0624 F/G 12/1 NL

END



NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

A Technical Report

Contract No. N00014-83-K-0624

CHARACTERIZATIONS OF GENERALIZED HYPEREXPONENTIAL DISTRIBUTIONS

Submitted to:

Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Attention: Group Leader, Statistics and Probability
Associate Director for Mathematical and Physical Sciences

Submitted by:

Robert F. Botta
Research Assistant

Carl M. Harris
Principal Investigator

Report No. UVA/525393/SE85/107

May 1985

SCHOOL OF ENGINEERING AND APPLIED SCIENCE

DEPARTMENT OF SYSTEMS ENGINEERING

UNIVERSITY OF VIRGINIA
CHARLOTTESVILLE, VIRGINIA 22901

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: UVA/525393/SE85/107	2. GOVT ACCESSION NO.: AD-A155 855
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Characterizations of Generalized Hyperexponential Distributions	5. TYPE OF REPORT & PERIOD COVERED: Technical Report
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Robert F. Botta and Carl M. Harris	8. CONTRACT OR GRANT NUMBER(s): N00014-83-K-0624
9. PERFORMING ORGANIZATION NAME AND ADDRESS: School of Engineering and Applied Science, Department of Systems Engineering, University of Virginia Charlottesville, VA 22901	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR 347-139
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Statistics & Probability Program Code 411SP	12. REPORT DATE: May 1985
	13. NUMBER OF PAGES: 42
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

probability distribution; cumulative distribution function; approximation; convergence in distribution; weak convergence; denseness; Erlang distribution; generalized hyperexponential distribution; method of stages

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Generalized hyperexponential (GH) distributions are linear combinations of exponential CDFs with mixing parameters (positive and negative) that sum to unity. The denseness of the class GH with respect to the class of all CDFs defined on $[0,\infty)$ is established by showing that a GH distribution can be found that is as close as desired, with respect to a suitably defined metric, to a given CDF. The metric induces the usual topology of weak convergence so that, equivalently, there exists a sequence $\{G_n\}$ of GH CDFs that converges

weakly to any CDF. The result follows from a similar well-known result for weak convergence of Erlang mixtures. Various set inclusion relations are also obtained relating the GH distributions to other commonly used classes of approximating distributions including generalized Erlang (GE), mixed generalized Erlang (MGE), those with reciprocal polynomial Laplace transforms (K_n), those with rational Laplace transforms (R_n), and phase-type (PH) distributions. A brief survey of the history and use of approximating distributions in queueing theory is also included.

A Technical Report

Contract No. N00014-83-K-0624

CHARACTERIZATIONS OF GENERALIZED HYPEREXPONENTIAL DISTRIBUTIONS

Submitted to:

Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Attention: Group Leader, Statistics and Probability
Associate Director for Mathematical and Physical Sciences

Submitted by:

Robert F. Botta
Research Assistant

Carl M. Harris
Principal Investigator

Department of Systems Engineering
SCHOOL OF ENGINEERING AND APPLIED SCIENCE
UNIVERSITY OF VIRGINIA
CHARLOTTESVILLE, VA 22901

Report No. UVA/525393/SE85/107 Copy No. ____

May 1985

ABSTRACT

Generalized hyperexponential (GH) distributions are linear combinations of exponential CDFs with mixing parameters (positive and negative) that sum to unity. The denseness of the class GH with respect to the class of all CDFs defined on $[0,\infty)$ is established by showing that a GH distribution can be found that is as close as desired, with respect to a suitably defined metric, to a given CDF. The metric induces the usual topology of weak convergence so that, equivalently, there exists a sequence $\{G_n\}$ of GH CDFs that converges weakly to any CDF. The result follows from a similar well-known result for weak convergence of Erlang mixtures. Various set inclusion relations are also obtained relating the GH distributions to other commonly used classes of approximating distributions including generalized Erlang, ~~(GE)~~, mixed generalized Erlang (MGE), those with reciprocal polynomial Laplace transforms (K_n), those with rational Laplace transforms (R_n), and phase-type ~~(PH)~~ distributions. A brief survey of the history and use of approximating distributions in queueing theory is also included.

Key phrases: probability distribution; cumulative distribution function; approximation; convergence in distribution; weak convergence; denseness; Erlang distribution; generalized hyperexponential distribution; method of stages.

1. INTRODUCTION

The purpose of this paper is to characterize the class of generalized hyperexponential (GH) probability distribution functions and to justify their use as convenient approximations to arbitrary CDFs.

1.1 Definition

Generalized hyperexponential distribution functions are of the form

$$F(t) = \sum_{i=1}^{n} a_i(1-e^{-\lambda_i t})$$

with $\sum_{i=1}^{n} a_i = 1$, a_i real, $\lambda_i > 0$. They are generalizations of the well-known hyperexponential distributions which are of the same form but with the additional requirement that the coefficients $\{a_i\}$ be positive. The familiar generalized Erlang CDFs arising as the distributions of a sum of independent, non-identical exponential random variables are in GH. A typical example is provided by the CDF

$$\begin{aligned} F(t) &= 3(1-e^{-t}) - 3(1-e^{-2t}) + (1-e^{-3t}) \\ &= 1 - 3e^{-t} + 3e^{-2t} - e^{-3t}. \end{aligned}$$

1.2 Organization

In the following, we first discuss briefly the evolution of approximations to CDFs in stochastic modeling, particularly in the field of queueing theory. Relationships among the classes of approximating distributions, including GH, are then developed in Section 2. Section 3 establishes that any CDF can be approximated as closely as desired, with

respect to a suitably defined metric, by a GH distribution. This fact, together with the attractive numerical and statistical properties of the class GH, provides a major justification for considering this class of approximants. Finally, Section 4 contains concluding remarks and some areas for future research.

1.3 Background

The use of approximating distributions in applied probability modeling dates back at least to the early part of the twentieth century. A. K. Erlang used the so-called method of stages to preserve the useful properties of exponential distribution functions in situations where the true underlying distributions were not in fact exponential (see, for example, Cox and Miller [1970]). By imagining customers in a queueing situation to progress through a series of independent stages in tandem, with the time spent in each stage having an exponential distribution, it is possible to preserve the Markovian character of the queueing system. The memoryless property of such systems simplifies the resulting equations governing queue behavior, such as the probability distributions of customer waiting time and number of customers in the system. Jensen [1954] generalized Erlang's technique, in part by allowing the exponential stages to have non-identical parameters.

Much of the queueing literature makes use of the theory of complex variables in the frequency domain which results when Laplace transforms of the probability distributions of interest are computed. Smith [1953] noted that the probabilities resulting from the method of stages have Laplace transforms that are reciprocal polynomials having negative real

roots. He extended the concept of stages by defining the class K_n to be all those distribution functions whose transforms are reciprocal polynomials of degree n with, in general, complex roots. He then showed, using Lindley's GI/G/1 formulation, that under mild conditions on the interarrival and service-time distributions, a service-time distribution of type K_n implies that the total equilibrium system time (queueing plus service) is also of type K_n. In particular, if service time is exponential, so is the system time for any distribution of interarrival times.

Cox [1955] extended the concept of stages further by considering the class of distributions having rational Laplace transforms. He showed that the method of stages can still be employed for this larger class of CDFs if one is willing to tolerate stages having complex roots and "probabilities" that may be negative. While the fictitious stages do not therefore correspond to physical entities, the resulting overall probabilities will be valid. The advantage of such an approach is that the desirable mathematical properties of Markovian systems may be retained. Cox went on to provide some justification for restricting attention to distributions with rational transforms by noting that if the degree of the polynomials is allowed to be countably infinite, any CDF can be closely approximated by one having a rational transform.

Wishart [1959] used the method of stages and Markov chains to verify Smith's K_n result for the equilibrium distribution of waiting times in a GI/G/1 queue having arbitrary interarrival-time distribution and service-time distribution characterized by a series of Erlang stages.

Kotiah et al. [1969] approximated the GI/G/1 queue by assuming that both the interarrival and service-time distributions were Erlangian, that is, consisted of a series of exponential stages. They developed numerical procedures to calculate the mean waiting time for the system and examined the effect of varying the skewness of the interarrival distribution.

Schassberger [1970] established the theoretical basis for some of the earlier work using the method of stages to obtain waiting-time distributions for the GI/G/1 queue. In doing so he showed how a sequence of mixtures of Erlang CDFs may be constructed that converge weakly to any desired distribution function defined on $[0,\infty)$.

Neuts [1975, 1981] has popularized the class of phase-type, or PH, probability distributions. These are distributions that arise or can be interpreted as the time until absorption in a finite Markov chain, and have rational Laplace transforms. Their major advantage is computational; instead of differential equations, complex variables and numerical integration, they admit of matrix-geometric procedures. A drawback of PH distributions, however, is the nonuniqueness of representation. Many different combinations of defining parameters lead to the same CDF and many of these representations are not of minimal order.

Theoretical justification for the use of approximating distributions has also been provided by work on the continuity of queues. Kennedy [1972, 1977] and Whitt [1974] have shown that if the interarrival and service-time distributions of otherwise identical queues are close in some sense, then the corresponding performance

measures such as queue length and waiting time will also be close in an appropriate sense. A very demanding technical treatment is needed to establish these results which requires careful definition of the underlying spaces, metrics, convergence concepts, and topologies. Both authors cite the sequence of mixed Erlang distributions, introduced by Schassberger that converges weakly to an arbitrary CDF. By constructing a sequence of such general Erlang models for a given GI/G/c queue, where the actual interarrival and service-time distributions are approximated, the weak convergence of the two sequences of CDFs implies the weak convergence of the corresponding performance measures.

This concept of weak convergence of probability measures has found widespread application in applied probability modeling. Queueing theory happens to be the area in which most of the weak convergence results have been used. Iglehart [1973] has written a useful survey paper that details the uses of weak convergence in queueing. Discussions on continuity of queues and rates of convergence are included.

Another interesting survey paper is that of Bhat et al. [1979]. They consider the use of approximations in queueing applications but their definition of approximation is somewhat broader than ours. Besides the use of approximating distributions, which they subsume under the heading of system approximations, they examine two other classes of approximations. Process approximations are concerned with replacing the physical process under study by a simpler one and include the use of diffusion and fluid approximations. Numerical approximation involves methods of simplifying the arithmetic computations that arise in solving the systems model; establishing upper and lower bounds on performance

measures and using numerical methods to invert analytically intractable Laplace transforms are examples of this type of approximation.

This concludes our brief review of the salient developments in the use of mixed-exponential-type approximations in applied probability. Although the emphasis has been on queueing applications, the basic concepts have wide applicability. While the family of mixed Erlang distributions has certainly been the most popular class of approximating functions, we will make a case in the sequel for considering the generalized hyperexponential distributions. Besides being of simple form which facilitates numerical manipulations, GH distributions have a unique representation which is desirable for such statistical procedures as parameter estimation. They extend the familiar hyperexponential class of distributions and enjoy the analytical benefits of having rational Laplace transforms. Furthermore, recently developed algorithms for fitting hyperexponential distributions to empirical data (see Kaylan and Harris [1981] and Mandelbaum and Harris [1982]) can be readily generalized to include GH distributions.

2. RELATIONS AMONG CLASSES OF DISTRIBUTION FUNCTIONS

In this section, families of probability distribution functions that find wide use as approximations to more general CDFs, for example, in queueing applications, are defined and related to one another. The more obvious relations are mentioned with the definitions, while others are presented in following subsections.

Several of the definitions below are stated in terms of the one-sided Laplace-Stieltjes transform of a CDF, F. This transform, F*, is defined in the ususal way as

$$F^*(s) = \int_0^\infty e^{-st}\, dF(t),$$

which is equivalent to the ordinary one-sided Laplace transform of a PDF, $F'(t) = f(t)$, whenever $F(t)$ is absolutely continuous.

2.1 Definitions

$\underline{K_n}$ Class

Smith [1953] defined the class K_n to be those distribution functions whose Laplace transform is the reciprocal of a polynomial of n^{th} degree. Of course, not all reciprocal polynomials are transforms of CDFs. For instance, the real part of each polynomial root must be negative. While the roots may be complex, they must occur in conjugate pairs since the corresponding CDF is real. There are also additional constraints that are not so obvious. Lukacs and Szasz [1951] have shown that one of the roots with greatest real part must be real. Therefore, the simplest member of K_n having complex roots is of the form

$$F^*(s) = \frac{a(a^2 + b^2)}{(s+a)[(s+a)^2 + b^2]}$$

distributions. For example, consider the two following distinct phase-type representations:

$$Q = \begin{bmatrix} -3 & 1 & 1 \\ 1 & -4 & 2 \\ 1 & 0 & -6 \end{bmatrix}, \qquad \underline{\alpha} = (0,\ 1/2,\ 1/2)$$

and

$$Q' = \begin{bmatrix} -2 & 0 \\ 0 & -5 \end{bmatrix}, \qquad \underline{\alpha}' = (2/3,\ 1/3) \ .$$

Clearly the two representations are different and are not of the same order. However, each results in the same CDF, namely, $F(t) = 1 - (2e^{-2t}/3 + e^{-5t}/3)$. The second representation would be of minimal order since the CDF is a mixture of two exponentials.

Mixed generalized Erlang distributions also permit multiple representations. From the notation of Dehon and Latouche [1982] we may represent the CDF of the sum of n independent random variables, each exponentially distributed with parameter λ_i $(i = 1,2,..,n)$, by $F_{12...n}$. This CDF is obtained in terms of the underlying exponentials by Equation (2.3.2). But the two CDFs defined by

$$F(t) = (1/3)\ F_1 + (2/3)\ F_{13}$$

and

$$G(t) = (1/3)\ F_1 + (4/9)\ F_{12} + (2/9)\ F_{123}$$

are in fact the same. This can be seen by expressing each as a linear combination of the underlying exponential distributions. As discussed above, this representation is unique and yields

$$F(t) = G(t) = (-1/3)\ F_1 + (4/3)\ F_3 \ .$$

2.6 Uniqueness of Representation

For statistical applications, an important property of mixture-type CDFs is uniqueness of representation, or identifiability. Yakowitz and Spragins [1968] define the identifiability of finite mixtures as follows. If $\{F_i\}$ is a collection of CDFs, then the class of finite mixtures of the $\{F_i\}$ is said to be identifiable if the convex hull of $\{F_i\}$ has the property that

$$\sum_{i=1}^{N} c_i F_i = \sum_{i=1}^{M} c_i' F_i'$$

where $c_i > 0$, $\Sigma c_i = 1$, implies $N = M$ and that for each i $(1 \le i \le N)$ there is some j $(1 \le j \le N)$ such that $c_i = c_j'$ and $F_i = F_j'$. A necessary and sufficient condition for identifiability is that the class $\{F_i\}$ be a linearly independent set over the field of real numbers. This follows from the uniqueness of representation property of a basis in a vector space.

Since any collection of distinct exponentials is linearly independent, the class of finite mixtures of exponential CDFs is identifiable. A broader concept of identifiability for generalized mixtures also applies when the underlying family of CDFs is exponential. A generalized mixture is one where the mixing parameters sum to unity but can have any real values, and of course, the GH distributions are of this form. Again, the uniqueness of the representation of vectors with respect to a basis for the vector space implies that GH distributions have unique representations as linear combinations of exponentials.

Importantly, the other families of CDFs considered in this work do not share the uniqueness of representation property with the GH

and MGE is a proper subset of PH. The results presented in Examples 2.3.1 and 2.4.1 are developed more fully in Botta [1985] where conditions are also given for GH and PH distributions (with real roots) to have MGE representations of the same order as the GH representation. These conditions are readily computed from the given distribution and do not require solving for the $\{b_i\}$ coefficients.

2.5 Summary of Set Inclusion Relations

The results of the foregoing subsections yield the following set of relations among the classes of distribution functions:

(1) $GE \subset K_n \subset R_n$

(2) $GE \subset MGE \subset GH \subset R_n$

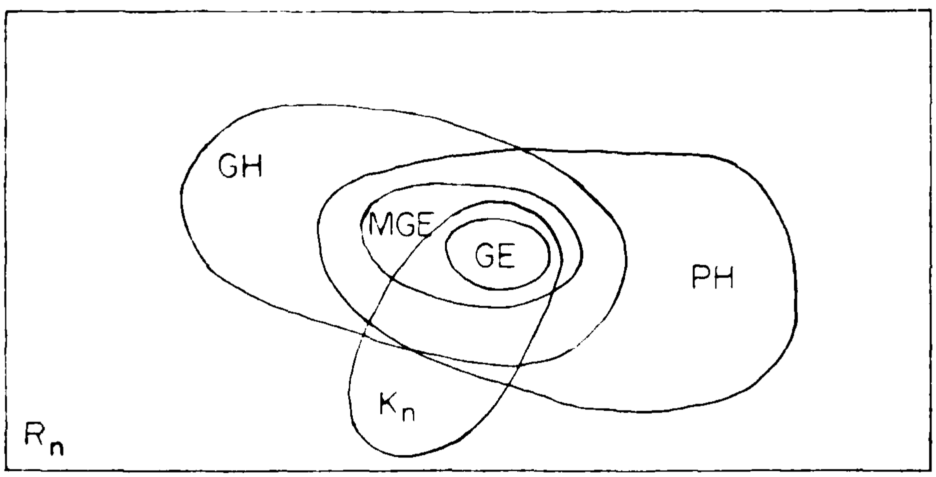
(3) $GE \subset MGE \subset PH \subset R_n$

(4) $PH \not\subset K_n$; $K_n \not\subset PH \Rightarrow R_n \not\subset PH$

(5) $PH \not\subset GH$; $GH \not\subset PH$

(6) $GH \not\subset MGE$ (of same order)

These relations can be depicted in the following Venn diagram.



in the subsection on uniqueness of representations, that it may be possible to obtain a MGE representation by embedding the problem in a higher order space even when there is no valid MGE representation in the original space.

2.4 MGE and PH

We established in subsection 2.1 that all MGE distributions are phase type. Since PH distributions may include trigonometric terms, it is clear that the MGE distributions are a proper subset of PH. But what if the PH generator matrix is allowed to have only real eigenvalues? Is the resulting subclass of PH distributions contained in MGE? The answer is no. We obtain this result by way of a counter example.

Example 2.4.1 The PH distribution given by

$$F(t) \equiv 1 - (1.293\, e^{-4.846t} - .343\, e^{-4.1948t} + .050\, e^{-.959t})$$

was obtained from the generator matrix

$$Q = \begin{bmatrix} -5 & 0 & 1/8 \\ 4 & -4 & 0 \\ 0 & 1 & -1 \end{bmatrix}$$

with $\underline{\alpha} = (1,0,0)$. As before, equating $F(t)$ to $b_1F_1(t) + b_2F_{12}(t) + b_3F_{123}(t)$ and solving for the $\{b_i\}$ yields the result that $b_2 = -.0369$. Since each b_i must be nonnegative, we do not have a valid MGE representation. Thus, PH distributions with real roots do not necessarily belong to MGE. In other words,

$$\text{PH (real roots)} \not\subset \text{MGE}$$

By substituting (2.3.2) in (2.3.3), a triangular system of linear equations relating the $\{a_i\}$ and $\{b_i\}$ coefficients is obtained. This system of equations is readily inverted to yield the $\{b_i\}$ in terms of the $\{a_i\}$. For the case of $n = 3$, it turns out that b_1 and b_3 are always nonnegative for any choice of $\{a_i\}$ corresponding to a GH distribution. The nonnegativity of b_2 requires that

$$a_2 \geq - \frac{\lambda_3(\lambda_1-\lambda_3)}{\lambda_2(\lambda_1-\lambda_2)} a_3 . \tag{2.3.4}$$

The next example shows that GH distributions exist for which (2.3.4) is violated.

Example 2.3.1 Consider the GH CDF

$$F(t) = 1 - (6e^{-4t} - 13e^{-3t} + 8e^{-2t}) .$$

Here

$$a_1 = 6,\ a_2 = -13,\ a_3 = 8$$
$$\lambda_1 = 4,\ \lambda_2 = 3,\ \lambda_3 = 2.$$

Therefore

$$- \frac{\lambda_3(\lambda_1-\lambda_3)}{\lambda_2(\lambda_1-\lambda_2)} a_3 = - \frac{32}{3} .$$

Since $a_2 < -32/3$, we see that (2.3.4) is violated and thus that no MGE representation exists for $F(t)$. This example establishes that

$$GH \not\subset MGE ,$$

and that the class of MGE distributions is thus a proper subset of the class of GH distributions.

The above result holds when the order of the MGE representation must be the same as that of the GH distribution. We demonstrate below,

where the A_i are real. Any mixture of such distributions has a transform of the same form. Therefore any mixed generalized Erlang distribution is in GH and

$$MGE \subset GH \quad . \tag{2.3.1}$$

Based upon results in Dehon and Latouche [1982], we next demonstrate the existence of GH distributions that cannot be represented as MGEs of the same order. They show that any GE distribution constructed from a subset of exponential distributions, $\{F_i\}$, can be expressed as a random combination of the GE distributions F_1, $F_{12}\ldots$, $F_{12\ldots n}$ where $F_{12\ldots i}$ is the distribution of the sum of the first i independent exponential random variables. Each such distribution function can be written as

$$F_{12..i}(t) = \sum_{j=1}^{i} \left\{ \prod_{\substack{k=1 \\ k\neq j}}^{i} \left(\frac{\lambda_k}{\lambda_h - \lambda_j}\right) F_j(t) \right\}, \qquad (t \geq 0) \tag{2.3.2}$$

where $F_j(t) = 1-e^{-\lambda_j t}$. (It has been assumed without loss of generality that $\lambda_1 > \lambda_2 >\ldots> \lambda_n$.) Since the $\{\lambda_j\}$ are constants, (2.3.2) is in the form of a GH distribution whose coefficients are determined by the $\{\lambda_j\}$, which agrees with (2.3.1). In order for a GH distribution,

$F(t) = 1 - \sum_{i=1}^{n} a_i e^{-\lambda_i t}$, to have a MGE representation, there must exist a set of nonnegative numbers $\{b_i, i = 1,2,\ldots,n\}$ which sum to one and satisfy the equation

$$1 - \sum_{i=1}^{n} a_i e^{-\lambda_i t} = \sum_{i=1}^{n} b_i \, F_{12\ldots i}(t) \quad . \tag{2.3.3}$$

Because of the trigonometric terms, F(t) is clearly not in GH. So

$$PH \not\subset GH \quad .$$

But does every GH distribution have a PH representation? The answer is no. As mentioned earlier, the density function corresponding to any PH distribution is strictly positive for all $t > 0$. The following example exhibits a GH distribution that violates this condition.

Example 2.2.2 Consider the GH distribution defined by

$$F(t) = 1-(4e^{-t} - 6e^{-2t} + 3e^{-3t})$$

with corresponding density

$$f(t) = F'(t) = 4e^{-t} - 12e^{-2t} + 9e^{-3t}$$

It can easily be shown that $f(t) = 0$ for both $t = 0$ and $t = \ln(3/2)$ and that $f(t) > 0$ for all other values of t. Therefore, $F(t) \in PH$ and

$$GH \not\subset PH \quad .$$

2.3 MGE and GH

Recall that the generalized Erlang (GE) distributions have Laplace transforms

$$\prod_{i=1}^{n} \frac{\lambda_i}{s+\lambda_i}$$

where the λ_i are distinct. Using a partial fraction expansion, this transform can be written as

$$\sum_{i=1}^{n} \frac{A_i}{s+\lambda_i}$$

boundary equation can be easily used to determine if a candidate exponential sum is in fact in GH. For sums of more than three exponential terms, the boundary equation could be determined in similar fashion but would be very involved and still not of much practical use in determining membership in GH.

We next develop some additional relations between the classes K_n, R_n, GE, MGE, PH, and GH.

2.2 <u>GH and PH</u>

From the preceding subsection we know that all PH distributions are in R_n. But if the roots of the denominator polynomial are complex, the corresponding distribution will not belong to GH. The following example displays such a PH distribution.

<u>Example 2.2.1</u> Consider the 3x3 generator matrix

$$Q = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 1 & 0 & -3 \end{bmatrix} .$$

The eigenvalues of Q, which are equal to the roots of the denominator polynomial of the Laplace transform of e^{Qt}, are

$$\lambda_1 = -.2307 \; ; \; \lambda_2, \lambda_3 = -2.8846 \pm .5897\, i$$

where $i = \sqrt{-1}$. The resulting PH distribution corresponding to an initial state vector $\underline{\alpha} = (1,0,0)$ is

$$F(t) = 1 - 1.1729\, e^{-.2307t}$$
$$- [.1729 \cos .5897t + .3868 \sin .5897t]\, e^{-2.8846t} .$$

Note that, unlike the usual hyperexponential distribution, we do not require that each a_i be nonnegative. This added freedom makes the GH distributions extremely versatile. Indeed, in the following section, we derive the critical characterization that any CDF on $[0,\infty)$ can be approximated as closely as desired with respect to an appropriate metric by a member of GH.

The Laplace transform of a GH distribution has the form

$$\sum_{i=1}^{n} \frac{a_i \lambda_i}{s+\lambda_i}$$

so we immediately note that

$$GH \subset R_n \quad . \tag{2.1.7}$$

Of course, not all linear combinations of exponentials of the form

$1 - \sum_{i=1}^{n} a_i e^{-\lambda_i t}$ with $\lambda_i > 0$ and $\sum_{i=1}^{n} a_i = 1$ are GH distributions.

For example, the monotonicity condition requires that $\sum_{i=1}^{n} a_i \lambda_i \geq 0$. Also, assuming λ_n to be the smallest of the λ_i, the corresponding coefficient a_n must be positive to insure proper asymptotic behavior as $t \to \infty$. Bartholomew [1969] has established a number of sufficient conditions for a linear combination of exponentials to be a GH distribution, but no simple set of conditions that are both necessary and sufficient is known. Dehon and Latouche [1982] have recently characterized the class of GH distributions by deriving a parametric equation of the boundary of the convex region constituting GH for the case n = 3. The geometric representation is obtained by choosing a set of basis vectors from the class of all GH distributions composed of linear combinations of three exponentials. It does not appear that the

yields rational expressions for each component of $V^*(s)$. Therefore, the probability distribution of each state belongs to R_n as does the distribution of the time until absorption. We have, therefore, the relation

$$PH \subset R_n \quad . \tag{2.1.6}$$

Phase-type distributions can easily be constructed with Laplace transforms which are not reciprocal polynomials, so that $PH \not\subset K_n$. But is it possible that every K_n distribution has a PH representation? The answer is no. Corollary 2.2.1 in Neuts [1981] establishes that any non-trivial PH distribution has a corresponding density function that is strictly positive for all $t > 0$. The PDF given earlier as (2.1.1) has a reciprocal-polynomial Laplace transform but the density function is zero wherever $\cos bt = 1$. Therefore, the corresponding distribution function is not in PH. We have then that $K_n \not\subset PH$ which implies that $R_n \not\subset PH$ and that PH is thus a proper subset of R_n.

It should be noted that, given an aribtrary CDF, there is no easy way to determine if it is in PH. One must search for a suitable generator matrix and set of initial conditions that will yield the desired distribution.

<u>GH Class</u>

The generalized hyperexponential distributions are CDFs of the form

$$1 - \sum_{i=1}^{n} a_i e^{-\lambda_i t}$$

with $\lambda_i > 0$ and real, $\sum_{i=1}^{n} a_i = 1$ and a_i real.

It should be noted that PH representations are not unique. That is, there may exist many different generator matrices of different orders that lead to the same CDF. Examples are given below in subsection 2.6. The problem of finding minimal representations of PH distributions, that is, where the order of Q is as small as possible, has not been solved. Neuts [1981] established that the class of PH distributions is closed under convolution and finite mixtures, though in general, infinite mixtures of PH distributions are not of phase type. However, if the mixing probabilities are discrete phase type, then the infinite mixture is also of phase-type.

From the preceding discussion it follows that MGE distributions are phase type, i.e.,

$$MGE \subset PH.$$

The representation (2.1.4) of a PH distribution was obtained from the distribution functions, $\underline{v}(t)$, of the individual states of the underlying Markov chain which are the solutions of

$$\frac{dv(t)}{dt} = \underline{v}(t) \cdot Q \qquad . \tag{2.1.5}$$

The solution to this equation is $\underline{v}(t) = \underline{v}(0)e^{Qt} = \underline{\alpha}e^{Qt}$. Taking the Laplace transform of (2.1.5) yields

$$sV^*(s) - \underline{v}(0) = V^*(s) \cdot Q \ ,$$

so that

$$V^*(s)\ (sI-Q) = \underline{v}(0) = \underline{\alpha}$$

or

$$V^*(s) = \underline{\alpha}\ (sI-Q)^{-1} \ .$$

Thus $(sI-Q)^{-1}$ is the Laplace transform of e^{Qt}, and each term in the inverse matrix of $sI-Q$ is a rational expression. Multiplication by $\underline{\alpha}$

$$Q = \begin{bmatrix} -q_{11} & q_{12} \cdots q_{1n} \\ q_{21} & -q_{22} \cdots q_{2n} \\ & \cdots\cdots \\ q_{n1} & q_{n2} \cdots -q_{nn} \end{bmatrix} \qquad \begin{array}{l} (q_{ii} > 0;\ q_{ij} \geq 0,\ i \neq j; \\ -q_{ii} + \sum_{i=1}^{n} q_{ij} \leq 0,\ i = 1,2,..,n). \end{array}$$

This generator matrix corresponds to an (n+1)-state Markov chain with state (n+1) being the absorbing barrier. The vector $\underline{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ is the vector of initial state probabilities at $t = 0$, and the vector $\underline{e}$ is an n-dimensional column vector of all ones. The entries, q_{ij}, in the generator matrix represent the instantaneous rate of the transition from state i to state j. Two examples of distribution functions with PH representations follow.

Example 2.1.1 The GE distribution of order n with parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$ has the representation $\underline{\alpha} = (1,0,0,..,0)$ and

$$Q = \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 \ldots\ldots\ldots\ldots 0 \\ 0 & -\lambda_2 & \lambda_2 \ldots\ldots\ldots\ldots 0 \\ 0 & 00 \ldots\ldots & -\lambda_{n-1} \quad \lambda_{n-1} \\ 0 & 00 \ldots\ldots 0 & -\lambda_n \end{bmatrix} .$$

Example 2.1.2 The mixed exponential distribution

$$F(t) = \sum_{i=1}^{n} \alpha_i (1 - e^{-\lambda_i t})$$

has the representation $\underline{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ and

$$Q = \begin{bmatrix} -\lambda_1 & 0 \ldots\ldots 0 \\ 0 & -\lambda_2 \ldots\ldots 0 \\ 0 & 0 \ \ldots \ -\lambda_n \end{bmatrix} .$$

When combined into a single fraction, this becomes the quotient of two polynomials, the degree of the denominator being n and the degree of the numerator n-1. This motivates the definition of R_n as the class of distributions whose transforms are rational, with n being both the degree of the denominator polynomial and the maximal degree of the numerator polynomial. We have therefore established that the class of mixed generalized Erlang distributions, denoted by MGE, is contained in R_n. Cox [1955] points out that both the convolution and the mixture of any pair of distributions in R_n yields another distribution with rational Laplace transform. Furthermore, all distributions in R_n are continuous except for possible atoms at the origin and the corresponding density function is positive everywhere in $(0,\infty)$ except at isolated points. Finally, it is obvious that

$$K_n \subset R_n \quad . \qquad (2.1.3)$$

PH Class

Neuts [1975, 1981] has popularized a class of distribution functions that he refers to as phase type, or PH, distributions. A CDF is said to be of phase type if it arises as the time until absorption in a finite-state continuous-time Markov chain. That is, F, is phase type if it can be written as

$$F(t) = 1 - \underline{\alpha} \cdot e^{Qt} \cdot \underline{e} \qquad (2.1.4)$$

where Q is the generator matrix and has the form

corresponding to the PDF

$$f(t) = ab^{-2}(a^2 + b^2)\, e^{-at} (1 - \cos bt) \quad (a > 0). \tag{2.1.1}$$

Clearly, the ordinary exponential distribution belongs to K_n. Since the Laplace transform of the distribution of a sum of independent random variables is the product of the Laplace transforms of their individual distributions, it follows that the generalized Erlang CDFs corresponding to a sum of independent, exponentially distributed random variables with distinct parameters are also in K_n. These generalized Erlangs, denoted GE, have transforms of the form

$$\prod_{i=1}^{n} \frac{\lambda_i}{s+\lambda_i} \qquad (\lambda_i > 0)$$

where $\lambda_i/(s+\lambda_i)$ is the transform of an exponential CDF having mean $1/\lambda_i$. If all the random variables are identically distributed, the resulting distribution is the (simple) Erlang of degree n, $E_n(\lambda)$, and its Laplace transform is just $\lambda^n/(s+\lambda)^n$. Therefore we see that $E_n(\lambda) \varepsilon K_n$ and

$$GE \subset K_n \tag{2.1.2}$$

<u>R_n Class</u>

While K_n contains GE, it does not contain mixtures of GE CDFs, i.e., distributions of the form $\sum_{i=1}^{n} a_i F_i$ with $a_i \geq 0$, $\sum_{i=1}^{n} a_i = 1$ and $F_i \varepsilon$ GE. For example, suppose each F_i is exponential. By the linearity of the Laplace transform, the transform of $\sum_{i=1}^{n} a_i F_i$ is

$$\sum_{i=1}^{n} a_i \frac{\lambda_i}{s+\lambda_i} .$$

As in the PH example, one of the MGE representations is not of minimal order.

For most applications, such as curve fitting, non-uniqueness of representation is a disadvantage. We now discuss a situation, mentioned in subsection 2.3, where obtaining a representation of non-minimal order may be useful. Suppose we have a GH distribution that does not have an MGE representation of minimal order. It may be possible to embed the distribution in a higher order space in such a way that an MGE representation is obtained. We illustrate the procedure via an example.

<u>Example 2.6.1</u> Consider the GH distribution

$$F(t) = 1 - \left(-\frac{13}{15} e^{-7t} + \frac{77}{12} e^{-4t} - \frac{35}{4} e^{-3t} + \frac{21}{5} e^{-2t}\right) .$$

Here $\lambda_1 = 7$, $\lambda_2 = 4$, $\lambda_3 = 3$, $\lambda_4 = 2$. Dehon and Latouche [1982] established that an MGE representation exists if, and only if, there exists a set of coefficients $\{b_i, i = 1,2,3,4\}$ such that

$$F(t) = b_1 F_1 + b_2 F_{12} + b_3 F_{123} + b_4 F_{1234}$$

with each b_i nonnegative and their sum being one. It can be shown that such a set of coefficients does not exist (b_3 is negative). Let us now add an additional exponential term, e^{-6t} and write

$$F(t) = 1 - \left(-\frac{13}{15} e^{-7t} + 0\, e^{-6t} + \frac{77}{12} e^{-4t} - \frac{35}{4} e^{-3t} + \frac{21}{5} e^{-2t}\right) .$$

Here, $\lambda_1' = 7$, $\lambda_2' = 6$, $\lambda_3' = 4$, $\lambda_4' = 3$, $\lambda_5' = 2$.

We must now solve for the coefficients $\{b_i'\}$ from

$$F(t) = b_1' F_1' + b_2' F_{12}' + b_3' F_{123}' + b_4' F_{1234}' + b_5' F_{12345}'$$

where the primes indicate that the corresponding terms are defined with respect to the $\{\lambda_i'\}$. It turns out that there is a solution for the $\{b_i'\}$ that results in the representation

$$F(t) = \frac{1}{4} F_1' + \frac{1}{3} F_{12}' + \frac{1}{24} F_{123}' + \frac{1}{24} F_{1234}' + \frac{1}{3} F_{12345}' .$$

Not only does this give us an MGE representation, it also confirms that the original F(t) is in fact a valid CDF since it can be expressed as a mixture of CDFs.

This example raises the question of whether it is possible to obtain an MGE representation for every GH distribution. The answer, of course, is no since all MGEs are of phase type and we have seen that there exist GHs that are not members of PH. A fuller discussion of the representation of GH distributions as MGEs, including a set of necessary and sufficient conditions that do not require solving for the $\{b_i\}$ coefficients, is contained in Botta [1985].

The uniqueness property provides a strong rationale for our interest in the GH class of distributions. We turn next to an examination of their suitability for providing approximations to arbitrary distribution functions.

3. DENSENESS RESULTS FOR GH DISTRIBUTIONS

In this section we establish a major justification for our interest in the class of generalized hyperexponential distributions by showing that GH CDFs are dense in the class of all cumulative distribution functions on the nonnegative real line. That is, any CDF can be approximated arbitrarily closely (with respect to some metric) by a member of GH. The result eventually follows from a similar result for Erlang mixtures (see, for example, Schassberger [1970], Whitt [1974], and Kennedy [1977]). A theorem from functional analysis concerning the approximation of a continuous function by an exponential sum is first extended to show that a certain class of probability density functions can be approximated by a GH density. Several intermediate results then lead to the desired denseness property of the class GH.

3.1 Denseness of Erlang Mixtures in the Topology of Weak Convergence

Consider an arbitrary CDF $F(t)$ on $[0,\infty)$. Define a sequence of general Erlang CDFs by

$$F_n(t) = F(0) + \sum_{k=1}^{\infty} [F(\tfrac{k}{n}) - F(\tfrac{k-1}{n})] \, E_n^k(t) \qquad (t \geq 0) \qquad (3.1.1)$$

where $E_n^k(t)$ is the k-fold convolution of the exponential CDF with mean $1/n$. Schassberger [1970], Whitt [1974], and Kennedy [1977] state that the sequence $\{F_n\}$ converges weakly to F. That is, $F_n(t)$ converges to $F(t)$ at each continuity point of F.

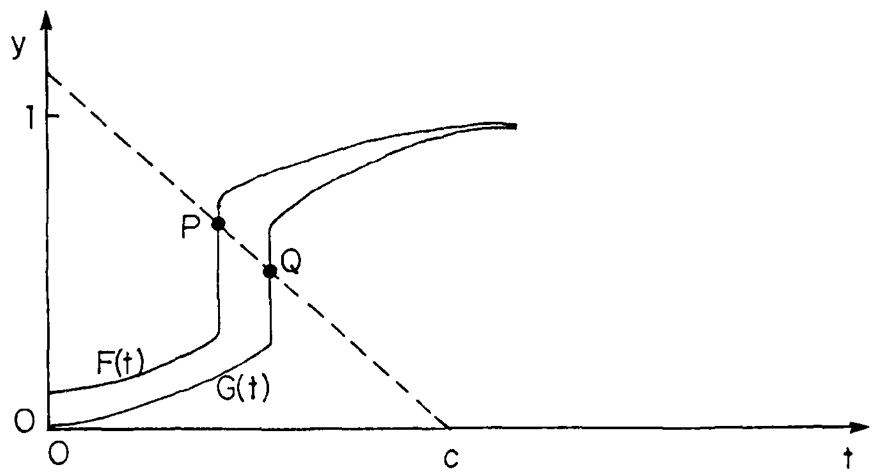
The notion of weak convergence induces a topology on the space of CDFs. The resulting topological space can also be generated by a number of metrics that measure the distance between any pair of CDFs. Convergence with respect to these metrics is then equivalent to

topological convergence. The resulting convergence in distribution, though weaker than the classical concepts of pointwise and uniform convergence, is useful for probabilistic modeling in situations where the stronger notions of convergence often fail. This occurs, for example, when the CDFs of interest have points of discontinuity.

A useful example of a metric defined on the space of CDFs is provided by the Levy distance. If F(t) and G(t) are two distribution functions, the Levy distance between them, denoted as L(F,G), is defined as

$$L(F,G) = \inf_{\varepsilon>0} \{\varepsilon \mid \text{for all } t,\ F(t-\varepsilon)-\varepsilon \le G(t) \le F(t+\varepsilon) + \varepsilon\}.$$

This analytic definition has an intuitive geometric interpretation. In the graphs of $y = F(t)$ and $y = G(t)$, vertical line segments are drawn at the points of discontinuity to produce two continuous curves. Let P and Q be the points on these curves that form the intersection of the curves with the line $t + y = c$. This is illustrated below.



Denoting by PQ the Euclidean distance between P and Q, the Levy distance can be expressed as

$$L(F,G) = \sup_{c} \frac{\overline{PQ}}{\sqrt{2}} \quad .$$

This definition illustrates that two CDFs can be close in the Levy sense if their points of discontinuity are close "horizontally" (i.e., $|t_1 - t_2|$ is small), even though they may not be close "vertically," that is, with respect to the usual sup metric which requires that $|F(t) - G(t)|$ be small for all values of t.

The connection between weak convergence and convergence with respect to the Levy metric is established by the following theorem from Lukacs [1975] which is stated here without proof. The geometric interpretation of L given above is from the same source and a proof of the theorem appears there as well.

<u>Theorem 3.1.1</u>: The sequence of CDFs $\{F_n(t)\}$ converges weakly to the CDF F(t) if, and only if, $\lim_{n\to\infty} L(F_n,F) = 0$.

It is important to note that the common statement that "a class of CDFs is dense in the class of all CDFs" generally is taken in the sense of the usual topology of weak convergence. That is the manner in which the Erlang mixtures of (3.1.1) are dense in the class of all CDFs with support on the nonnegative real line.

3.2 Approximating with Exponential Sums

In this subsection we establish that a continuous function on $[0,\infty)$ that vanishes at infinity can be uniformly approximated by a sum of exponential terms of the form

$$\Sigma\, a_i e^{-\lambda_i t} \qquad (\lambda_i > 0) \quad .$$

The result follows from the extension to an infinite domain of the famous Weierstrass polynomial approximation theorem. We present first the case where the λ_i are integers and then a generalization to arbitrary λ_i. The following lemma from Apostol [1974] is stated without proof.

<u>Lemma 3.2.1</u> If f is continuous on $[0,\infty)$ and if $f(t) \to a$ as $t \to \infty$, then f can be uniformly approximated on $[0,\infty)$ by a function of the form $g(t) = p(e^{-t})$, where p is a polynomial.

We now show that if the continuous function being approximated vanishes at infinity, the constant term in the approximating exponential sum can be set equal to zero.

<u>Lemma 3.2.2</u> If f is continuous on $[0,\infty)$ and if $f(t) \to 0$ as $t \to \infty$, then f can be uniformly approximated on $[0,\infty)$ by an exponential sum of the form

$$\sum_{k=1}^{n} a_k e^{-kt}.$$

<u>Proof</u>: By Lemma 3.2.1, f can be uniformly approximated by the sum of the form

$$a_o + \sum_{k=1}^{n} a_k e^{-kt}.$$

Thus we have only to show that a_o may be chosen to be zero. For $\varepsilon > 0$, let

$$\hat{f}_\varepsilon = a_o(\varepsilon) + \sum_{k=1}^{n(\varepsilon)} a_k(\varepsilon)\, e^{-kt}$$

uniformly approximate f, that is, $|f - \hat{f}_\varepsilon| \le \varepsilon$ for all $t \in [0,\infty)$.

Now consider

$$|a_o(\varepsilon)| = |\hat{f}_\varepsilon - \sum_{k=1}^{n(\varepsilon)} a_k(\varepsilon)e^{-kt}| = |\hat{f}_\varepsilon - f + f - \sum_{k=1}^{n(\varepsilon)} a_k(\varepsilon)e^{-kt}|.$$

Thus

$$|a_o(\varepsilon)| \le |\hat{f}_\varepsilon - f| + |f| + |\sum_{k=1}^{n(\varepsilon)} a_k(\varepsilon)e^{-kt}|.$$

But $\lim_{t\to\infty} f(t) = 0$ and clearly $\lim_{t\to\infty} \sum_{k=1}^{n(\varepsilon)} a_k(\varepsilon)e^{-kt} = 0$.

Therefore, for any $\alpha > 0$ there exists a value T such that $t > T$ implies

that $|f(t)| \le \alpha$ and $|\sum_{k=1}^{n(\varepsilon)} a_k(\varepsilon)\, e^{-kt}| \le \alpha$. We then have

$$|a_o(\varepsilon)| \le |\hat{f}_\varepsilon - f| + 2\alpha \le \varepsilon + 2\alpha .$$

Since α was arbitrary, it follows that

$$|a_o(\varepsilon)| \le \varepsilon.$$

But now consider the modified approximant

$$\bar{f} = \hat{f}_\varepsilon - a_o(\varepsilon) = \sum_{k=1}^{n(\varepsilon)} a_k(\varepsilon)e^{-kt}.$$

For any value of t

$$|f - \bar{f}| = |f - \hat{f}_\varepsilon + a_o(\varepsilon)| \le |f - \hat{f}_\varepsilon| + |a_o(\varepsilon)| \le 2\varepsilon.$$

Since ε is arbitrary, a uniform exponential approximation to f having a zero constant term can always be found.

Q.E.D.

We now state without proof a generalization of this result that permits the coefficients of t in the exponents of the approximating function to be non-integer. The lemma is found in Kammler [1976] and is based upon the Muntz-Szasz theorem (see Cheney [1966]).

Lemma 3.2.3 Let $0 < \lambda_1 < \lambda_2 < \ldots$ and assume that $\sum_{i=1}^{\infty} (1/\lambda_i)$ diverges. Then the set of exponential sums that may be written as finite linear combinations of the functions $e^{-\lambda_i t}$, $i = 1,2,\ldots$, is dense in the space of continuous functions on $[0,\infty)$ that vanish at infinity. In other words, a continuous function on $[0,\infty)$ that vanishes at infinity can be uniformly approximated by a linear combination of exponentials where the coefficients of t in the exponents need not be integers.

3.3 Approximating PDFs with Exponential Sums

We wish to develop an exponential sum approximation to a probability density function. For a particular class of PDFs -- those whose tails decay at least exponentially fast -- the results of the preceding section can be applied to show that the class GH is dense with respect to the PDFs of interest. That is, we approximate a PDF with an exponential sum that is also a PDF.

Theorem 3.3.1 Let f be a PDF continuous on $[0,\infty)$ and let $f \to 0$ exponentially fast as $t \to \infty$. That is, $\lim_{t\to\infty} f(t)e^{\lambda_o t} = 0$ for some $\lambda_o > 0$. Then f can be uniformly approximated on $[0,\infty)$ by a generalized hyperexponential PDF.

Proof: The proof consists of three parts. First we find an exponential sum approximation; next, we modify the approximation so that it is nonnegative; finally, we normalize the approximation so that its area is unity.

(i) Let $g(t) = f(t)\, e^{\lambda_o t}$. By Lemma 3.2.3 we can approximate $g(t)$ by a function of the form

$$\hat{g} = \sum_{k=1}^{n} a_k e^{-\lambda_k t} \qquad (\lambda_k > 0)$$

such that $|g - \hat{g}| \le \varepsilon$ for all $t \in [0,\infty)$. Thus we may write

$$|g - \hat{g}| = \left| f(t)e^{\lambda_o t} - e^{\lambda_o t} \sum a_k e^{-(\lambda_o+\lambda_k)t} \right| \le \varepsilon$$

or

$$e^{\lambda_o t} \left| f(t) - \sum a_k e^{-(\lambda_o+\lambda_k)t} \right| \le \varepsilon .$$

Therefore

$$\left| f(t) - \sum a_k e^{-(\lambda_o+\lambda_k)t} \right| \le \varepsilon e^{-\lambda_o t} \le \varepsilon . \tag{3.3.1}$$

This shows that $\hat{f} \equiv \sum a_k e^{-(\lambda_o+\lambda_k)t}$ uniformly approximates f. Of course, $\hat{f}$ may be negative for some values of t and so may not be a valid PDF. (More on this subsequently.)

(ii) From (3.3.1) we have

$$|f(t) - \hat{f}(t)| \le \varepsilon e^{-\lambda_o t} \tag{3.3.2}$$

so that $0 \le f(t) \le \hat{f}(t) + \varepsilon e^{-\lambda_o t}$, where the first inequality follows from the fact that f is a PDF. Define the right-hand side to be

$$\bar{f} \equiv \hat{f}(t) + \varepsilon e^{-\lambda_o t} \ge f(t) \ge 0 . \tag{3.3.3}$$

Then

$$|f - \bar{f}| = |f - \hat{f} - \varepsilon e^{-\lambda_o t}| \le |f - \hat{f}| + \varepsilon e^{-\lambda_o t}$$

or

$$|f - \bar{f}| \le \varepsilon e^{-\lambda_o t} + \varepsilon e^{-\lambda_o t} = 2\varepsilon e^{-\lambda_o t} \le 2\varepsilon . \tag{3.3.4}$$

Therefore, $\bar{f}$ is a nonnegative exponential sum that uniformly approximates f. However, $\bar{f} \geq f$ from (3.3.3), so that

$$\int_0^\infty \bar{f}\, dt \geq \int_0^\infty f\, dt = 1$$

and $\bar{f}$ may not be a PDF.

(iii) To produce an approximation to f that is indeed a PDF, we must normalize f so that its area is unity. Let

$$A = \int_0^\infty \bar{f}\, dt \geq 1.$$

If A = 1, then f is a PDF and we are finished. If A > 1, define

$f' = \bar{f}/A$, so that $\int_0^\infty f' dt = 1$. It remains to show that f' uniformly

approximates f on $[0,\infty)$. From (3.3.2) we have

$$\hat{f}(t) \leq f(t) + \varepsilon e^{-\lambda_o t} .$$

Using (3.3.3)

$$\bar{f}(t) = \hat{f}(t) + \varepsilon e^{-\lambda_o t} \leq f(t) + 2\varepsilon e^{-\lambda_o t} .$$

Therefore

$$A = \int_0^\infty \bar{f}\, dt \leq \int_0^\infty f\, dt + \int_0^\infty 2\varepsilon e^{-\lambda_o t}\, dt = 1 + \frac{2\varepsilon}{\lambda_o} . \qquad (3.3.5)$$

Now consider

$$|f - f'| = |f - \frac{1}{A}\bar{f}| = \frac{1}{A}|Af - \bar{f}|$$

$$= \frac{1}{A}|Af - f + f - \bar{f}| = \frac{1}{A}|(A-1)f + f - \bar{f}|$$

$$\leq \frac{A-1}{A}|f| + \frac{1}{A}|f - \bar{f}| \leq (A-1)|f| + |f - \bar{f}| .$$

The last inequality follows from (3.3.5). Finally, from (3.3.4) and (3.3.5), we obtain

$$|f-f'| \leq \frac{2\varepsilon}{\lambda_o} |f| + 2\varepsilon \leq \frac{2\varepsilon}{\lambda_o} \|f\| + 2\varepsilon. \tag{3.3.6}$$

The second of these inequalities follows from the boundedness of f, which in turn is a consequence of the continuity of f and the fact that $f \to 0$ as $t \to \infty$ (see, for example, Boas [1972], p. 78). Since the RHS of (3.3.6) can be made as small as desired by an appropriate choice of ε, f' uniformly approximates f, is nonnegative, and integrates to unity and therefore is a valid PDF. Furthermore,

$$f' = \frac{\varepsilon}{A} e^{-\lambda_o t} + \sum_{k=1}^{n} \frac{a_k}{A} e^{-(\lambda_o+\lambda_k)t} = \sum_{k=0}^{n} \frac{a_k}{A} e^{-\lambda_k t} \quad (\lambda_k > 0)$$

where $a_o = \varepsilon$. Therefore, $f' \varepsilon$ GH.

Q.E.D.

Let us now consider the class R_n of PDFs having rational Laplace transforms, where n is the degree of the denominator polynomial. The roots of the denominator each have negative real part so that when a partial fraction expansion is formed and the inverse transform taken, there are at most n terms, each of the form $t^k e^{-\alpha t}$ (A cos bt + B sin bt). Therefore, the PDF goes to zero exponentially fast and is continuous. In other words, all PDFs that are in R_n satisfy the conditions of Theorem 3.3.1. We have then the following corollary.

Corollary: Every PDF in R_n can be uniformly approximated on $[0,\infty)$ by a generalized hyperexponential density. That is, GH PDFs are dense in R_n.

3.4 Approximating CDFs with Exponential Sums

In this subsection we wish to extend the exponential sum approximation to cumulative distribution functions (CDFs). We begin by showing that if two PDFs are close in some sense, then their corresponding CDFs are also close. It then follows that any finite mixture of Erlang CDFs can be approximated by a generalized hyperexponential CDF. The results of subsection 3.1 are then used to show that any CDF can be closely approximated by a generalized hyperexponential CDF.

Lemma 3.4.1. Let f be a PDF continuous on $[0,\infty)$. If another PDF, g, uniformly approximates f, then the CDF $G = \int_0^t g(x)\,dx$ uniformly approximates the CDF $F = \int_0^t f(x)\,dx$ on $[0,\infty)$.

Proof: For any $\varepsilon > 0$ there exists a value t_o such that for $t \geq t_o$, $F(t) \geq 1 - \frac{\varepsilon}{2}$. This follows from the existence of the integral $\int_0^\infty f(x)\,dx = F(\infty) = 1$ by the Cauchy criterion (see, for example, Bartle [1964], p. 345). Let g be such that $|f - g| \leq \varepsilon/2t_o$ for all $t \in [0,\infty)$ where, for the moment, we assume $t_o \neq 0$. We now examine $|F - G|$ on the intervals $[0,t_o]$ and $[t_o,\infty)$.

(i) $[0,t_o]$

$$|F - G| = \left|\int_0^t f\,dx - \int_0^t g\,dx\right| = \left|\int_0^t (f - g)\,dx\right|$$

$$\leq \int_0^t |f - g|\,dx \leq \int_0^{t_o} |f - g|\,dx \leq \frac{\varepsilon}{2}\,.$$

(ii) $[t_o,\infty)$

From (i), $|F(t_o) - G(t_o)| \leq \varepsilon/2$, so that $G(t_o) \geq F(t_o) - \varepsilon/2 \geq 1 - \varepsilon/2 - \varepsilon/2 = 1 - \varepsilon$. By the monotonicity of G it follows that $G(t) \geq G(t_o) \geq 1 - \varepsilon$ for all $t \geq t_o$. Therefore, on $[t_o, \infty)$, $F - G \geq 1 - \varepsilon/2 - G \geq - \varepsilon/2$ since $G(t) \leq 1$ for all t. Also $F - G \leq 1 - G \leq 1-(1 - \varepsilon) = \varepsilon$. Therefore, $|F - G| \leq \varepsilon$. Combining the results from (i) and (ii) we have that $|F - G| \leq \varepsilon$ on $[0,\infty)$, so that G uniformly approximates F.

The only way that t_o could be zero is if $\varepsilon/2 \geq 1$. However, $|F - G| \leq |F| + |G| \leq 1 + 1 = 2 \leq \varepsilon$; so again G uniformly approximates F.

Q.E.D.

At this point, we pause to note that we have established the desired denseness property of the class GH with respect to a subset of CDFs. In particular, if F is an absolutely continuous CDF on $[0,\infty)$ and its derivative is continuous and has an exponentially decaying tail, then it follows from Theorem 3.3.1 and Lemma 3.4.1 that there exists a GH CDF that uniformly approximates F. In other words, we can find a $G \in GH$ with the property that $|F(t) - G(t)| < \varepsilon$ for all $t \in [0,\infty)$.

Continuing with our general development, we note that an Erlang PDF is defined on $[0,\infty)$ and has a Laplace transform of the form $(\lambda/(\lambda + s))^n$ where λ is a positive real number. Consequently the Erlang PDFs belong to R_n and, from the corollary to Theorem 3.3.1, we obtain the following corollary to the preceding lemma.

Corollary: Every Erlang CDF can be uniformly approximated on $[0,\infty)$ by a GH CDF.

Recall that $E_n^k(t)$ is the Erlang CDF obtained by taking the k-fold convolution of the exponential CDF $1-e^{-nt}$. Let us use the notation $G_n^k(t)$ to represent a GH CDF that uniformly approximates $E_n^k(t)$ on $[0,\infty)$. We now use the result stated in subsection 3.1 to show that any CDF on $[0,\infty)$ can be approximated arbitrarily closely by a generalized hyperexponential CDF.

Theorem 3.4.1 Let F be an arbitrary CDF defined on $[0,\infty)$. Then a generalized hyperexponential CDF can be found that approximates F arbitrarily closely in the topology of weak convergence. In other words, the set of generalized hyperexponential CDFs is dense in the set of all CDFs defined on $[0,\infty)$.

Proof: From Equation (3.1.1) the sequence of CDFs defined by

$$F_n = F(0) + \sum_{k=1}^{\infty} [F(\tfrac{k}{n}) - F(\tfrac{k-1}{n})]\, E_n^k(t) \qquad (3.4.1)$$

converges to F at each continuity point of F. By the corollary to Lemma 3.4.1, there exists a GH distribution that uniformly approximates E_n^k on $[0,\infty)$, call it G_n^k. Therefore

$$|E_n^k - G_n^k| \le \varepsilon \quad \text{on } [0,\infty). \qquad (3.4.2)$$

Let $F(\frac{k}{n}) - F(\frac{k-1}{n}) = b_n^k$ and define H_n as

$$H_n = F(0) + \sum_{k=1}^{\infty} b_n^k \; G_n^k \; . \tag{3.4.3}$$

The existence of H_n can be characterized as follows. Since G_n^k is a CDF it never exceeds unity. Therefore,

$$\sum_{k=1}^{\infty} b_n^k \; G_n^k \leq \sum_{k=1}^{\infty} b_n^k = 1 - F(0)$$

by the definition of b_n^k. Since both G_n^k and b_n^k are nonnegative, the sequence of partial sums

$$\sum_{k=1}^{K} b_n^k \; G_n^k \qquad .$$

is bounded above and monotonically increasing with K, and so it has a limit. At each continuity point t of F we have that $\lim_{n\to\infty} F_n(t) = F(t)$. That is, for $\varepsilon > 0$ there exists an $N(\varepsilon,t)$ such that for all $n \geq N$, $|F_n(t) - F(t)| \leq \varepsilon$. We are now ready to show that $H_n(t)$ approximates $F(t)$.

$$\begin{aligned} |H_n(t) - F(t)| &= |H_n(t) - F_n(t) + F_n(t) - F(t)| \\ &\leq |H_n(t) - F_n(t)| + |F_n(t) - F(t)| \\ &\leq |H_n(t) - F_n(t)| + \varepsilon . \end{aligned} \tag{3.4.4}$$

From Equations (3.4.1) and (3.4.3),

$$\begin{aligned} |H_n(t) - F_n(t)| &= \left| \sum_{k=1}^{\infty} b_n^k \left(G_n^k(t) - E_n^k(t)\right) \right| \\ &\leq \sum_{k=1}^{\infty} b_n^k \; |G_n^k(t) - E_n^k(t)| . \end{aligned}$$

By Inequality (3.4.2), this becomes

$$|H_n(t) - F_n(t)| \le \varepsilon \sum_{k=1}^{\infty} b_n^k \le \varepsilon .$$

Substituting in (3.4.4) yields

$$|H_n(t) - F(t)| \le 2\varepsilon, \quad n \ge N(\varepsilon, t). \tag{3.4.5}$$

Since ε is arbitrary, for every value of t the sequence of CDFs $\{H_n\}$ approximates F as closely as desired. Each approximant, $H_n(t)$, where n depends upon t and ε, consists of an infinite sum of GH CDFs. We now show that the infinite sum may be replaced by a finite sum.

It follows from the definition of b_n^k that there exists a number $K^*(n)$ such that for all $K \ge K^*(n)$,

$$\sum_{k=K}^{\infty} b_n^k \le \frac{1}{n} .$$

Now define

$$H_n^{K^*(n)} = F(0) + \sum_{k=1}^{K^*(n)-1} b_n^k \; G_n^k(t) + \sum_{k=K^*(n)}^{\infty} b_n^k . \tag{3.4.6}$$

Next, consider the sequence of functions $\{H_n^{K^*(n)}\}$. For each $\varepsilon > 0$, there exists $N(\varepsilon, t)$ such that for all $n \ge N$, $|H_n(t) - F(t)| \le \varepsilon$ by (3.4.5). Now choose $n^*(\varepsilon, t) = \max(N, 1/\varepsilon)$. Therefore, for all $n \ge n^*$ we have

$$\begin{aligned} |H_n^{K^*(n)}(t)| &= |H_n^{K^*(n)}(t) - H_n(t) + H_n(t) - F(t)| \\ &\le |H_n^{K^*(n)}(t) - H_n(t)| + |H_n(t) - F(t)| \qquad (3.4.7) \\ &\le |H_n^{K^*(n)}(t) - H_n(t)| + \varepsilon . \end{aligned}$$

The last inequality holds since $n \geq n^* \geq N$. Now from (3.4.3) and (3.4.6),

$$|H_n^{K^*(n)}(t) - H_n(t)| = \left| \sum_{k=K^*(n)}^{\infty} b_n^k G_n^k(t) - \sum_{k=K^*(n)}^{\infty} b_n^k \right|$$

$$= \left| \sum_{k=K^*(n)}^{\infty} b_n^k (G_n^k(t) - 1) \right|$$

$$\leq \sum_{k=K^*(n)}^{\infty} b_n^k \leq \frac{1}{n} \leq \varepsilon$$

since $|G_n^k(t) - 1| \leq 1$ and $n \geq n^* \geq 1/\varepsilon$. Substituting (3.4.8) into (3.4.7) yields

$$|H_n^{K^*(n)}(t) - F(t)| \leq \varepsilon + \varepsilon = 2\varepsilon, \quad n \geq n^*. \tag{3.4.9}$$

By the way $H_n^{K^*(n)}$ was constructed, it is a CDF and (3.4.9) establishes that $\{H_n^{K^*(n)}\}$ converges weakly to F. Each $H_n^{K^*(n)}(t)$ contains a finite linear combination of CDFs each of which is GH. In the event that $F(0) = 0$, $H_n^{K^*(n)}$ is a (finite) convex combination of these GH CDFs and so is itself GH. When $F(0) > 0$, we can write $H_n^{K^*(n)}$ as the mixture

$$H_n^{K^*(n)}(t) = p_1\, U(t) + p_2 \left[\sum_{k=1}^{K^*(n)-1} b_n^k\, G_n^k(t) \Big/ \sum_{k=1}^{K^*(n)-1} b_n^k \right]$$

where

$$p_1 = F(0) + \sum_{k=K^*}^{\infty} b_n^k\,, \quad p_2 = \sum_{k=1}^{K^*-1} b_n^k$$

and $U(t) \equiv 1$ is the CDF of an atom at $t = 0$. From the definition of the $\{b_n^k\}$, $p_1 + p_2 = 1$. If the atom at $t = 0$ is thought of as an exponential distribution with vanishingly small mean, $H_n^{K^*(n)}$ can be viewed as a GH CDF for any value of $F(0)$.

To recapitulate, we have demonstrated the existence of a sequence of GH CDFs, $\{H_n^{K^*(n)}\}$, that converges to a given CDF, F, at each of its continuity points.

Q.E.D.

If the limiting CDF is continuous, then weak convergence becomes pointwise convergence. A result due to Polya, cited on p. 86 of Chung [1974], establishes that the convergence is in fact uniform in this case. Therefore, any continuous CDF with support on the nonnegative real line can be uniformly approximated by GH CDFs.

4. CONCLUDING REMARKS

We have made a case for considering generalized exponential mixtures to approximate any CDF defined on $[0,\infty)$ by demonstrating that the class GH is dense in the class of all CDFs, i.e., any CDF can be approximated as closely as desired by a member of GH. Therefore, GH joins other known dense classes of probability distributions such as those of phase-type and those having rational Laplace transforms. In addition to the denseness property, GH distributions have a unique representation; this property is not shared by all dense classes of distributions. We also presented a set of relations positioning the GH class among other often used classes of distribution functions. The properties of the GH class of distributions make it attractive for both numerical and statistical computations.

This work has focused on theoretical results and does not discuss the important area of how to construct an approximating GH distribution. Recent work, however, has extended to generalized exponential mixtures a maximum likelihood-based algorithm for fitting mixed Weibull distributions to empirical data. Questions that remain for future investigation include determining the number of terms required for a finite mixture to be "good enough" and the related question of the minimum achievable distance between a given CDF and the class of GH distributions having a fixed number of terms.

REFERENCES

Apostol, T. M. (1974). Mathematical Analysis, 2nd ed. Reading, Mass.: Addison-Wesley.

Bartholomew, D. J. (1969). Sufficient Conditions for a Mixture of Exponentials to be a Probability Density Function. Annals of Mathematical Statistics, 40, 2183-2188.

Bartle, R. G. (1964). The Elements of Real Analysis. New York: Wiley.

Bhat, U. N., Shalaby, M. and Fischer, M. J. (1979). Approximation Techniques in the Solution of Queueing Problems. Naval Research Logistics Quarterly, 26, 311-326.

Boas, R. P. (1972). A Primer of Real Functions, 2nd ed. The Carus Mathematical Monographs. The Mathematical Association of America, No. 13.

Botta, R. F. (1985). Approximation of Probability Distribution Functions by Generalized Exponential Mixtures. Unpublished Ph.D. Dissertation, Department of Systems Engineering, University of Virginia, Charlottesville, VA.

Cheney, E. W. (1966). Introduction to Approximation Theory. New York: McGraw-Hill.

Chung, K. L. (1974). A Course in Probability Theory, 2nd ed. New York: Academic.

Cox, D. R. (1955). A Use of Complex Probabilities in the Theory of Stochastic Processes. Proceedings of the Cambridge Philosophical Society, 51, 313-319.

Cox, D. R. and Miller, H. D. (1970). The Theory of Stochastic Processes. London: Methuen.

Dehon, M. and Latouche, G. (1982). A Geometric Interpretation of the Relations Between the Exponential and Generalized Erlang Distributions. Advances in Applied Probability, 14, 885-897.

Harris, C. M. and Sykes, E. A. (1985). Likelihood Estimation for Generalized Mixed Exponential Distributions. Operations Research, to appear.

Iglehart, D. L. (1973). Weak Convergence in Queueing Theory. Advances in Applied Probability, 5, 570-594.

Jensen, A. (1954). A Distribution Model Applicable to Economics. Copenhagen: Munksgaard.

Kammler, D. W. (1976). Approximations with Sums of Exponentials in $L_p[0,\infty)$. Journal of Approximation Theory, 16, 384-408.

Kaylan, A. R. and Harris, C. M. (1981). Efficient Algorithms to Derive Maximum-Likelihood Estimates for Finite Exponential and Weibull Mixtures. Computers and Operations Research, 8, 97-104.

Kennedy, D. P. (1972). The Continuity of the Single Server Queue. Journal of Applied Probability, 9, 370-381.

Kennedy, D. P. (1977). The Stability of Queueing Systems. Bulletin of the International Institute of Statistics, 47, 355-365.

Kotiah, T. C. T., Thompson, J. W., and Waugh, W. A. O. (1969). Use of Erlangian Distributions for Single-Server Queueing Systems. Journal of Applied Probability, 6, 584-593.

Lukacs, E. (1975). Stochastic Convergence, 2nd ed. New York: Academic.

Lukacs, E. and Szasz, O. (1951). Certain Fourier Transforms of Distributions. Canadian Journal of Mathematics, 3, 140-144.

Mandelbaum, J. and Harris, C. M. (1982). Parameter Estimation under Progressive Censoring Conditions for a Finite Mixture of Weibull Distributions. In TIMS Studies in the Management Sciences 19. Amsterdam: North Holland, 239-260.

Natanson, I. P. (1964). Theory of Functions of a Real Variable, Vol. I, revised ed. (translated from the Russian). New York: Ungar.

Neuts, M. F. (1975). Probability Distributions of Phase Type. In Liber Amicorum Professor Emeritus Dr. H. Florin. Belgium: University of Louvain, 173-206.

Neuts, M. F. (1981). Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach. Baltimore: Johns Hopkins University Press.

Schassberger, R. (1970). On the Waiting Time in the Queueing System GI/G/1. Annals of Mathematical Statistics, 41, 182-187.

Smith, W. L. (1953). On the Distribution of Queueing Times. Proceedings of the Cambridge Philosophical Society, 49, 449-461.

Whitt, W. (1974). The Continuity of Queues. Advances in Applied Probability, 6 175-183.

Wishart, D. M. G. (1959). A Queueing System with Service Time Distribution of Mixed Chi-Squared Type. Operations Research, 7, 174-179.

Yakowitz, S. J. and Spragins, J. D. (1968). On the Identifiability of Finite Mixtures. Annals of Mathematical Statistics, 39, 209-214.

DISTRIBUTION LIST

Copy No.	
1 - 20	Office of Naval Research 800 North Quincy Street Arlington, VA 22217 Attention: Program Manager Statistics and Probability
21	ONR Resident Representative Joseph Henry Building Room 623 2100 Pennsylvania Avenue, N.W. Washington, D.C. 20037
22	Defense Technical Information Center Building 5, Cameron Station Alexandria, VA 22314
23 - 24	C. M. Harris
25 - 26	R. F. Botta
27	C. C. White
28 - 29	E. H. Pancake Sci./Tech. Information Center
30	SEAS Publications Files

JO# 5817:lsl

UNIVERSITY OF VIRGINIA
School of Engineering and Applied Science

The University of Virginia's School of Engineering and Applied Science has an undergraduate enrollment of approximately 1,500 students with a graduate enrollment of approximately 500. There are 125 faculty members, a majority of whom conduct research in addition to teaching.

Research is a vital part of the educational program and interests parallel academic specialties. These range from the classical engineering disciplines of Chemical, Civil, Electrical, and Mechanical and Aerospace to newer, more specialized fields of Biomedical Engineering, Systems Engineering, Materials Science, Nuclear Engineering and Engineering Physics, Applied Mathematics and Computer Science. Within these disciplines there are well equipped laboratories for conducting highly specialized research. All departments offer the doctorate; Biomedical and Materials Science grant only graduate degrees. In addition, courses in the humanities are offered within the School.

The University of Virginia (which includes approximately 1,500 full-time faculty and a total full-time student enrollment of about 16,000), also offers professional degrees under the schools of Architecture, Law, Medicine, Nursing, Commerce, Business Administration, and Education. In addition, the College of Arts and Sciences houses departments of Mathematics, Physics, Chemistry and others relevant to the engineering research program. The School of Engineering and Applied Science is an integral part of this University community which provides opportunities for interdisciplinary work in pursuit of the basic goals of education, research, and public service.

END

FILMED

8-85

DTIC